

# Netflix Rating Prediction

Shupeng Niu : sniu  
Kartikeya Pharasi : kpharas  
Alok Kucheria : akucher

October 15, 2015

## 1 Data Set

The Netflix Prize Dataset consists of multiple files with attributes such as movie id, title, user id, rating and dates.

## 2 Project Ideas

First, we will find the clusters of similar movies from public sources based on genre. Then we will try to cluster users based on the movie watching habits. Based on data from the clusters, we can calculate various factors including number of ratings, average ratings, variance of ratings.

Combining these factors and using classifiers, we can predict the rating a user will give on a movie from the movies that user has rated, as well as the ratings similar users have given similar movies.

## 3 Software

We will use R to build a classifier to train and test the dataset.

## 4 Work Division

Shupeng Niu: Coding, Data Preprocessing  
Kartikeya Pharasi: Coding, Function Design  
Alok Kucheria: Coding, Classifier Selection

## 5 Midterm milestone

- 1.Data Preprocessing
- 2.Clustering movies and users.
- 3.Identify all the functions needed for classification.
- 4.Choose part of the training data and test data, test our algorithm and modify it if needed.

## References

- [1]Koren, Yehuda, Robert Bell, and Chris Volinsky. *Matrix factorization techniques for recommender systems*. Computer 8 (2009): 30-37.
- [2]Cremonesi, Paolo, Yehuda Koren, and Roberto Turrin. *Performance of recommender algorithms on top-n recommendation tasks*. Proceedings of the fourth ACM conference on Recommender systems. ACM, 2010.